

Multi-Model Hypothesize-and-Verify Approach for Incremental Loop Closure Verification

Tanaka Kanji

Abstract—Loop closure detection, which is the task of identifying locations revisited by a robot in a sequence of odometry and perceptual observations, is typically formulated as a visual place recognition (VPR) task. However, even state-of-the-art VPR techniques generate a considerable number of false positives as a result of confusing visual features and perceptual aliasing. In this paper, we propose a robust incremental framework for loop closure detection, termed incremental loop closure verification. Our approach reformulates the problem of loop closure detection as an instance of a multi-model hypothesize-and-verify framework, in which multiple loop closure hypotheses are generated and verified in terms of the consistency between loop closure hypotheses and VPR constraints at multiple viewpoints along the robot's trajectory. Furthermore, we consider the general incremental setting of loop closure detection, in which the system must update both the set of VPR constraints and that of loop closure hypotheses when new constraints or hypotheses arrive during robot navigation. Experimental results using a stereo SLAM system and DCNN features and visual odometry validate effectiveness of the proposed approach.

I. INTRODUCTION

Loop closure detection, which is the task of identifying locations revisited by a robot in a sequence of odometry and perceptual observations, is a major first step to robotic mapping, localization and SLAM [1]. Failure in loop closure detection can yield catastrophic damage in an estimated robot trajectory, and achieving an acceptable tradeoff between precision and recall is critical in this context. Loop closure detection is typically formulated as a visual place recognition (VPR) task. However, even state-of-the-art VPR techniques generate a considerable number of false positives as a result of confusing visual features and perceptual aliasing.

In this study, we propose a robust incremental framework for loop closure detection, which we call incremental loop closure verification. Our approach reformulates loop closure detection as an instance of a multi-model hypothesize-and-verify problem in which a set of hypotheses of robot trajectories is generated from VPR constraints using a general pose graph SLAM [2], and verified for the consistency between loop closure hypotheses and VPR constraints at multiple viewpoints along the robot's trajectory. Furthermore, we consider the general incremental setting of loop closure detection, in which the system must update both the set of VPR constraints and that of loop closure hypotheses when new constraints or hypotheses arrive during robot navigation.

Our work has been supported in part by JSPS KAKENHI Grant-in-Aid for Young Scientists (B) 23700229, and for Scientific Research (C) 26330297.

K. Tanaka is with Graduate School of Engineering, University of Fukui, Japan. tnkknj@u-fukui.ac.jp

The proposed approach is motivated by three independent observations. First, we are inspired by the recent success of hypothesize-and-verify techniques (e.g., USAC [3]). Second, loop closure detection is essentially a multi-model estimation problem [4], rather than a single model estimation considered in classical applications of the hypothesize-and-verify approach (e.g., structure-from-motion [3]), where the goal is to identify multiple instances of models (i.e., loop closure hypotheses) and where the inliers to one model behave as pseudo-outliers to the other models. Finally, and most importantly, the framework is sufficiently general and effective for implementing various hypothesize-and-verify strategies that implement various types of domain knowledge.

Although the proposed approach is general, we focus on a challenging SLAM scenario to demonstrate the efficacy of the proposed system. Our experiments employ a stereo SLAM system that implements stereo visual odometry as in [5], loop closure detection using appearance-based image

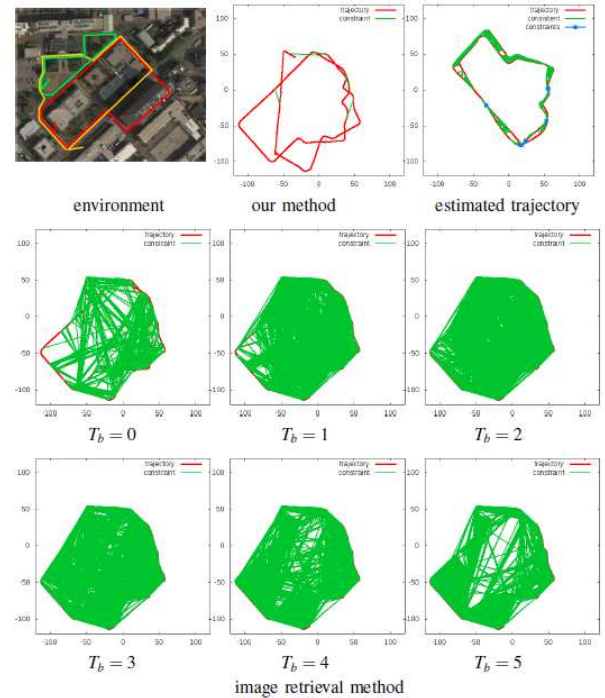


Fig. 1. Loop closure detection (1st row) Left: environment with robot trajectories. Red, yellow, and green lines indicate viewpoint paths on which datasets 1-3 were collected. Middle: four loops detected by our method. Right: trajectory estimated by loop closing using the detected loops as constraints. (2nd, 3rd rows) Loop closure constraints detected by using information from only the image retrieval. T_b is the threshold on the dissimilarity metric (i.e., Hamming distance) between visual features (i.e., binary codes from DCNN features) employed by the image retrieval system.

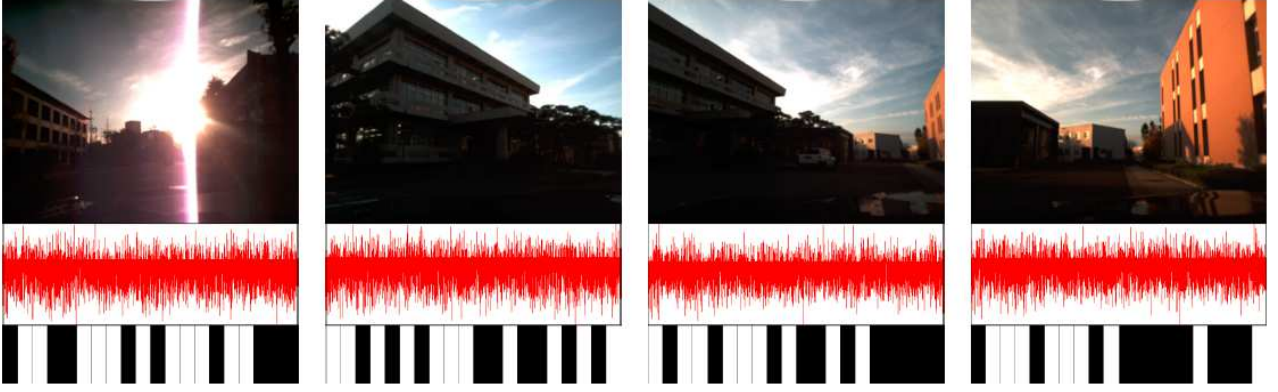


Fig. 2. Examples of visual features used by the image retrieval system. For each figure, the top, middle and bottom panels are input image, the 4096-dimensional DCNN feature and its binary code.

retrieval with DCNN features as in [6] and binary landmark as in [7], and finally, pose graph SLAM as in [2]. Fig. 1 illustrates an odometry-based robot trajectory, as well as a trajectory corrected by loop closing, and a set of VPR constraints selected by our method. As can be seen, major errors in trajectory are accumulated as the robot navigates and these errors are corrected given correct VPR constraints. Fig. 1 also reveals that image retrieval-based loop closure detection is less than perfect; a considerable number of false positives and negatives are present. These two types of errors, (i.e., accumulated errors in odometry and misrecognition in VPR of errors), are the main error sources that we address in this study. Experimental results using our stereo SLAM system confirm that the proposed strategy, which includes the use of VPR constraints and loop closure hypotheses as a guide, achieves promising results despite the fact that many false positive constraints and hypotheses exist.

The proposed approach is orthogonal to many existing approaches to loop closure detection. In literature, most of the existing works on loop closure detection have focused on the image retrieval step in the task, rather than the verification step [1]. In fact, loop closure detection techniques are typically classified in terms of image retrieval strategies (rather than post-verification strategies) [8]. Images are typically represented by a collection of invariant local descriptors [6] or a global holistic descriptor [9]. Loop closure detection has been employed by many SLAM systems [10]. However, the above works did not focus on the post-verification step or introduce novel insight to the hypothesize-and-verify framework.

This paper is a part of our studies on loop closure detection. In [11], we addressed the issue of guided sampling strategies for loop closure verification, by which the number of image matching over all the locations is minimized, whereas the current study assumes a constant number of image matching at each location and focuses on the hypothesize-and-verify approach to loop closure verification. The proposed is partially inspired by an incremental extension of multi-modal hypothesize-and-verify approach in [12], where a different task of robot relocation was considered. Recently, we have discussed robust VPR algo-

rithms [13], landmark discovery [14], and DCNN landmarks [15] in IROS15, ICRA15 and IROS16 papers. Hypothesize-and-verify strategy in loop closure detection has not been addressed in the above papers.

II. APPROACH

A. Loop Closure Detection

For clarity of presentation, we first describe a baseline SLAM system in which the proposed approach is built and used as a benchmark for performance comparison in the experimental section. As previously mentioned, we build the proposed system on a stereo SLAM system in which a stereo vision sensor is employed for both visual odometry [5] and visual feature acquisition [6]. In addition, we follow the standard formulation of pose graph SLAM [2]. In pose graph SLAM, the robot is assumed to move in an unknown environment, along a trajectory described by a sequence of variables $x_{1:T} = x_1, \dots, x_T$. While moving, it acquires sequences of odometry measurements $u_{1:T} = u_1, \dots, u_T$ and perceptual measurements $z_{1:T} = z_1, \dots, z_T$. Each odometry measurement u_t ($1 \leq t \leq T$) is a pairing of rotation and translation acquired by visual odometry. Each perceptual measurement z_t is both a set of VPR constraints $z_t^1, \dots, z_t^{N_r}$, and a pair of location IDs, t, t' . This measurement also has a similarity score that represents the likelihood that the location pair belongs to the same place. In our case, this score is obtained by the VPR task. More formally, we begin with an empty history of VPR constraints. At each time t , we run the VPR using the latest visual image as a query to identify the top $N_r = 10$ ranked retrieved images that obtain the highest similarity scores. We then insert the N_r pairs from the query image and each of the N_r top-ranked retrieved images as new constraints to the history. To prevent trivial VPR constraints, images are acquired when the interval $[t - \Delta t, t + \Delta t]$ is not considered a candidate for the VPR constraint ($\Delta t = 200$).

For simplicity, we begin by assuming that fixed sets of VPR constraints $z_{1:T}$ and N_h trajectory hypotheses $m_{1:M}$ are a priori given. Typical hypothesize-and-verify algorithms require such a fixed set assumption [4]. Clearly, this assumption is violated in our SLAM applications as both the

VPR constraints and trajectory hypotheses must be incrementally derived as the robot navigates. This incremental setting is addressed in Section II-D by relaxing the fixed set assumption. We divide the entire measurement sequence into constant time windows and generate $N_g = 100$ new hypotheses per window. To generate a hypothesis, we employ pose graph SLAM that expects the following as input: 1) an existing trajectory hypothesis and 2) a new VPR constraint selected from the history $z_{1:T}$ of loop closure constraints. In experiments, we set the time window size to $W = 500$ frames.

The performance is evaluated in terms of quality of estimated trajectory. As previously mentioned, a trajectory hypothesis can be obtained by performing the pose graph SLAM using the loop closure hypothesis as input. To evaluate the quality of a given trajectory hypothesis, we first compute a set of VPR constraints that are consistent with the hypothesis, count the numbers of true positives N_{TP} , false positives N_{FP} , and false negatives N_{FN} . We then evaluate the precision and recall in the form of $N_{TP}/(N_{TP} + N_{FP})$ and $N_{TP}/(N_{TP} + N_{FN})$, respectively. This performance measure requires a set of ground truth VPR constraints. For each query image i , we define pairs of locations (t, t') that satisfy

$$\|p(t, h) - p(t', h)\| < T_p \quad (1)$$

as ground-truth VPR constraints. In (1), $p(t, h)$ is the two-dimensional coordinate of location t conditioned on a robot trajectory hypothesis h , and T_p is the preset threshold of 10 m.

B. VPR Constraints

The image retrieval system encodes the image to a DCNN feature representation as in [6]. First, we extract a 4,096 dimensional DCNN feature from an image. Although a DCNN is composed of several layers in each of them responses from the previous layer are convoluted and activated by a differentiable function. We use the sixth layer of DCNN, because it has proven to produce effective features with excellent descriptive power in previous studies [6]. We then perform PCA compression to obtain 128 dimensional features. Our strategy is supported by the recent findings in [6] in which PCA compression provides excellent short codes with 128 short vectors that generate state-of-the-art accuracy on several recognition tasks. In our experiments, we use DCNN features from the image collection to train PCA models for different settings of the output dimension of 128. However, direct use of DCNN features for image retrieval is computationally demanding as it requires many-to-many comparisons of high-dimensional DCNN features between the query and the image collection. To address this concern, we employ a compact binary encoding of images and fast bit-count operation that enables fast image comparison (Fig. 2). Query and library features are encoded to $N_b = 20$ -bit binary codes using the compact projection technique borrowed from [16] and then compared using the Hamming distance to obtain a set of candidate images. The, L2 distance of high-dimensional DCNN features between the query and each

element of the set are then computed and the top- N_r elements with the lowest L2 distance are inserted into the constraint history $z_{1:T}$. In this study, the threshold T_b for the Hamming distance and the parameter N_r are empirically set to 3 and 10, respectively. Fig.2 shows several examples of input images, DCNN features and binary codes.

C. Hypothesization

At each iteration, the system generates a set of N_h hypotheses from N_h constraints that are selected from the constraint history $z_{1:T}$. The task of selecting N_h constraints can be formulated as an iterative process of selecting the i -th constraint given a sequence $C^{(i-1)}$ of $(i-1)$ constraints chosen thus far. Therefore, the remaining problem is the manner by which select the i -th constraint. To address this issue, we present a simple strategy using the trajectory hypothesis.

Our approach begins with an empty sequence $C^{(0)}$ of constraints. It randomly selects the first constraint $z^{(1)}$ from the constraint history $z_{1:T}$. We then run the pose graph SLAM using the selected VPR constraint to generate a trajectory hypothesis $h^{(1)}$ that is consistent with $z^{(1)}$. Intuitively, the next constraint $z^{(2)}$ should be *inconsistent* with the trajectory hypothesis $h^{(1)}$, because we want to obtain a new trajectory hypothesis $h^{(2)}$ that is dissimilar from the existing hypothesis $h^{(1)}$. In general, the i -th constraint should be inconsistent with the trajectory hypothesis $h^{(i-1)}$. To implement this idea, we propose to select the next constraint $z^{(i)}$ from those constraints $\{(t, t')\}$ whose distance exceeds a pre-defined threshold:

$$\|p(t, h^{(i-1)}) - p(t', h^{(i-1)})\| > T_p. \quad (2)$$

D. Incremental Extension

In this section, we relax the fixed set assumption given in Section II-A and consider the general incremental setting of loop closure detection, in which the system must update the sets of VPR constraints and loop closure hypotheses. Most part of the proposed algorithm work properly with the incremental setting.

One exception is that the hypothesis list and constraint history are no longer fixed but grow linear to the time. The main space cost for a hypothesis includes 1) IDs of VPR constraints that are used for generating the hypothesis, and 2) estimated the trajectory or a sequence of estimated robot poses in 3DOF, which is linear to time. Because the number of hypotheses is also linear to time, the total space cost grows quadratically with time. Fortunately, in this study, the length of VPR constraints is shorter than 2000 and the number of hypotheses is smaller than 200. As a result, the number of evaluated hypothesis-constraint pairs is small. Specifically, it is less than $2000 \times 200 = 400,000$ per iteration. Furthermore, we observe that one can eliminate the lowest ranked hypotheses to save space and time costs, as their contribution to the overall performance is typically minimal. Time cost is nearly constant. Although strictly speaking, the evaluation of hypotheses using constraints grows quadratically with time,

this evaluation can be performed quickly using a look-up table that stores the results of image retrieval (i.e., indicating which constraint is consistent with which hypothesis). This look-up table must be updated when new hypotheses and constraints arrive, a process that requires a constant cost per time.

Another exception is with respect to the hypothesis generation procedure. For each time window, we iterate the hypothesis generation step for 10 times. Then, in each hypothesis generation step, 10 hypotheses are generated from 10 top-ranked hypotheses. This yields $N_g = 10 \times 10$ new hypotheses per time window.

Fig.3 and 4 show examples of incremental hypothesis generation.



Fig. 3. Iterative process of selecting constraints. Each r -th row corresponds to the hypothesis that received an r -th rank at the end of the navigation. For each row, from left to right, each i -th panel shows the trajectory estimated by using $1, \dots, i-1$ constraints selected.

III. EXPERIMENTS

A. Settings

We conducted loop closure detection experiments using a stereo SLAM system in a university campus. Our experiments employed a stereo SLAM system that implemented the proposed strategies for loop closure detection. The main steps involved visual odometry, loop closure detection, and post-verification. The first step executed stereo visual odometry in order to reconstruct the robot trajectory. We adopted the stereo visual odometry algorithm proposed in [5], which has proven to be effective in recent visual odometry applications (e.g., [17]). The second step applied the appearance-based image retrieval with DCNN features, a fast visual search using a binary landmark [7] and precise image matching using DCNN features [6]. This second step generates a set of new N_r VPR constraints and inserted these into the history of VPR constraints as mentioned in Section II-B. The third step performed incremental loop closure verification to generate and verify hypotheses in terms of the consistency between loop closure hypotheses and VPR constraints from multiple viewpoints along the robot's trajectory. This step also incorporates new hypotheses and constraints from the

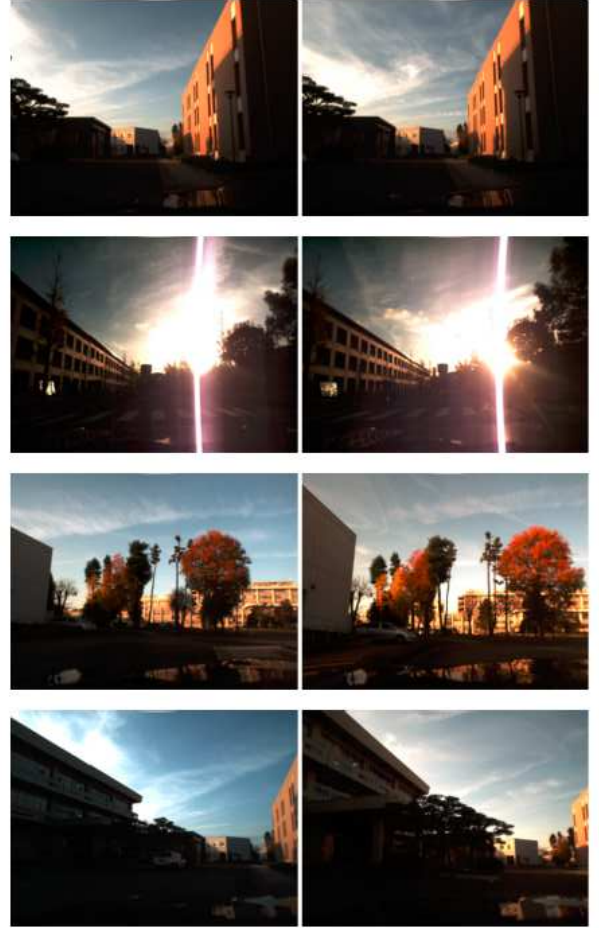


Fig. 4. Loop closure constraints detected by the proposed method. Left: query image. Right: retrieved image.

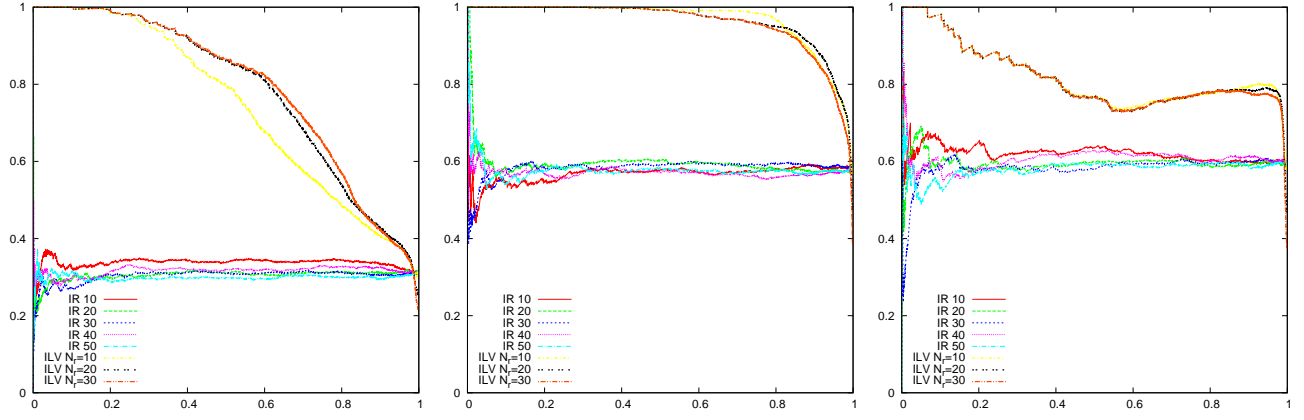


Fig. 5. Precision recall curves. Vertical and horizontal axes indicate precision and recall, respectively. IR N_r : image retrieval only. N_r loop closure constraints randomly selected from image retrieval are used to estimate the trajectory. ILV N_r : proposed method using a different setting for N_r . Top N_r -ranked images from the image retrieval are considered as new N_r VPR constraints.

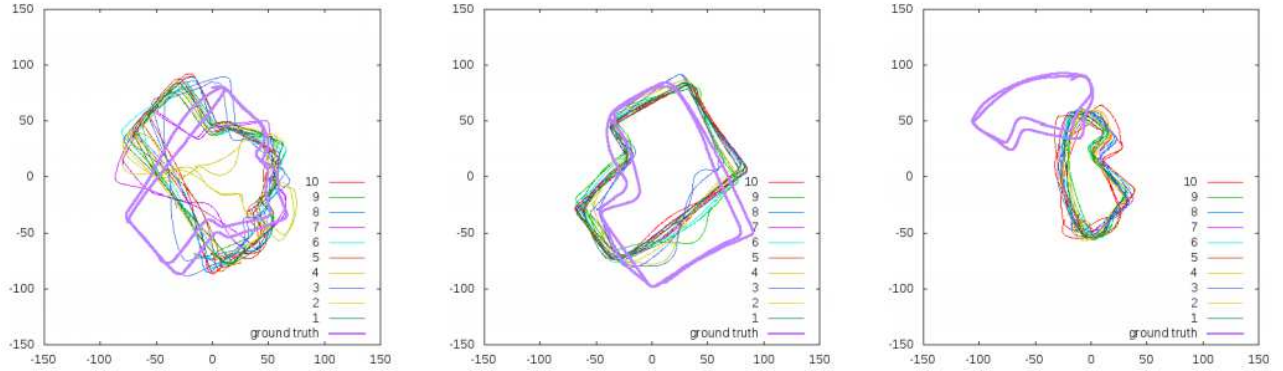


Fig. 6. Ten trajectories estimated by using 10 hypotheses that received top-10 ranks at the end of navigation. Each colored line corresponds to the i -th ranked hypothesis. The thick line shows the ground truth trajectory.

VPR task.

Fig. 1 presents robot trajectories superimposed on Google map imagery. The ground-truth trajectories were generated using a SLAM algorithm based on the graph optimization in [2] using manually identified ground-truth VPR constraints as input. As indicated, major odometry errors were collected as the robot navigated. We collected three sequences along routes with travel distances of 756, 657, and 401 m, respectively, using a cart equipped with a Bumblebee stereo vision camera system, as illustrated in Fig.4. We defined a ground-truth loop closure constraint as a pairing of two locations i, j whose distance was less than 10 m. Occlusion was severe in the scenes and people and vehicles were dynamic entities occupying the scenes. We processed each path and collected three stereo image sequences with lengths of 1651, 1621, and 979, respectively. We used images with a size of 640×480 pixels from the left-eye view of the stereo camera as input for the image retrieval system.

B. Results

Fig.5 shows the performance of loop closure detection in terms of estimated trajectory accuracy. We can see that the proposed method (ILV) achieves good tradeoff between

precision and recall. No clear correlation exists between the parameter N_r and the performance of the proposed method. The figure also shows results for loop closure detection using only image retrieval (i.e., combining the binary landmark [7] and DCNN feature [6]). In this case, we randomly sampled X ($X \in \{10, 20, 30, 40, 50\}$) VPR constraints from the history of VPR constraints and used the samples to estimate the trajectory using the pose graph SLAM. Fig.5 reveals that these methods (i.e., loop closing with image retrieval only) do not achieve high precision performance, which indicates the effectiveness of the proposed post-verification framework.

Fig.6 provides examples of estimated trajectories for 10 hypotheses that were top-ranked at the end of the navigation. We can see that most of the top-ranked hypotheses are successful at estimating sufficiently accurate trajectories.

Fig.7 visualizes consistency between hypotheses and constraints for three time windows of 1000, 1500, and 2000. This figure reveals that the number of hypotheses is linear to time. As expected, top-ranked hypotheses (e.g., hypotheses ranked 0-10 as shown in the right panels) are consistent with a greater number of constraints than are those hypotheses shown in the figure.

Fig.8 shows examples of the sequential hypothesis gener-

ation.

IV. CONCLUSIONS

The main contribution of this study is a novel robust framework for loop closure detection, termed incremental loop closure verification. Our approach reformulated the problem of loop closure detection as an instance of a multi-model hypothesize-and-verify framework, in which multiple loop closure hypotheses are generated and verified using VPR results at multiple viewpoints along the robot's trajectory. Then, we considered the general incremental setting of loop closure detection, where the system must update the set of VPR constraints and set of loop closure hypotheses when new constraint or hypothesis arrives during the robot navigation. Experimental results using a stereo SLAM system and DCNN features and visual odometry validated effectiveness of the proposed approach.

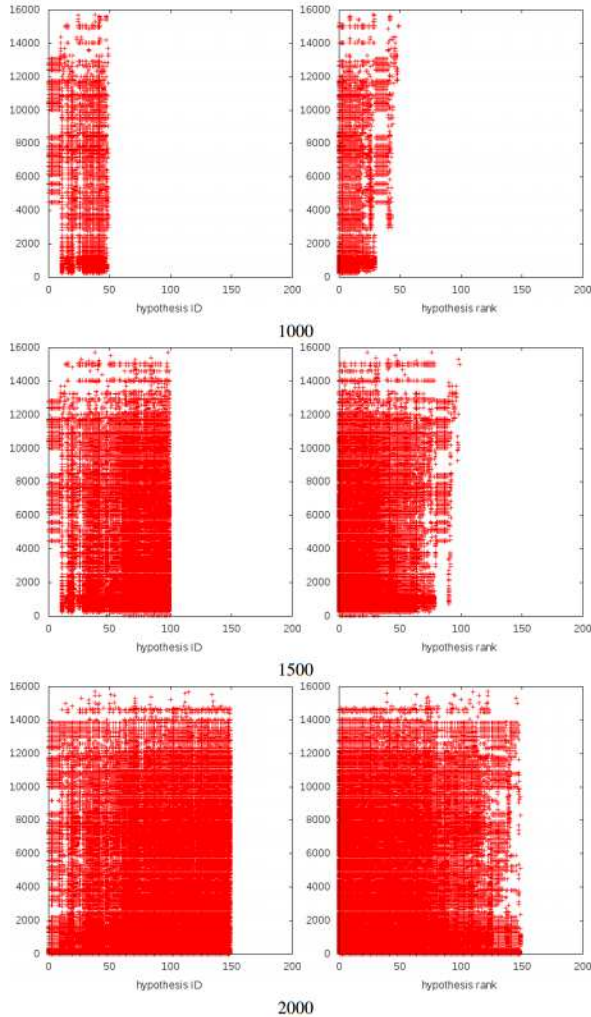


Fig. 7. Consistent hypothesis-constraint pairs for different time windows of 1000, 1500 and 2000. Vertical axis: constraint ID. Horizontal axis: hypothesis ID (left) and hypothesis rank (right).

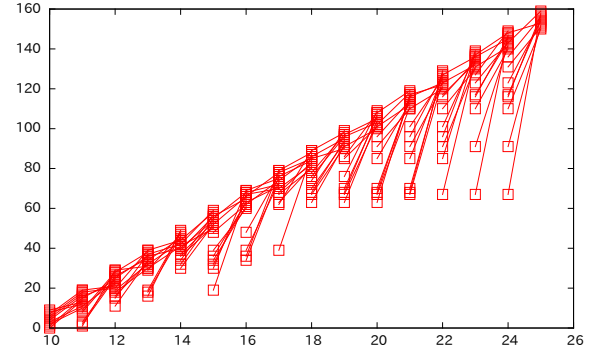


Fig. 8. Sequential hypothesis generation. Vertical axis: hypothesis ID. Horizontal axis: iteration ID. Each line connects one of the top-10 ranked hypotheses and to the next generation hypothesis generated from it.

REFERENCES

- [1] B. P. Williams, M. Cummins, J. Neira, P. M. Newman, I. D. Reid, and J. D. Tardós, "A comparison of loop closing techniques in monocular SLAM," *Robotics and Autonomous Systems*, vol. 57, no. 12, pp. 1188–1197, 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.robot.2009.06.010>
- [2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.
- [3] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. Frahm, "Usac: a universal framework for random sample consensus," *IEEE T. PAMI*, vol. 35, no. 8, pp. 2022–2038, 2013.
- [4] Y. Kanazawa and H. Kawakami, "Detection of planar regions with uncalibrated stereo using distributions of feature points," in *BMVC*. Citeseer, 2004, pp. 1–10.
- [5] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV)*, 2011 IEEE. IEEE, 2011, pp. 963–968.
- [6] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," pp. 584–599, 2014.
- [7] K. Ikeda and K. Tanaka, "Visual robot localization using compact binary landmarks," in *Robotics and Automation (ICRA)*, 2010 IEEE International Conference on. IEEE, 2010, pp. 4397–4403.
- [8] B. P. Williams, M. Cummins, J. Neira, P. M. Newman, I. D. Reid, and J. D. Tardós, "An image-to-map loop closing method for monocular SLAM," in *IROS*, 2008, pp. 2053–2059.
- [9] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE T. Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.
- [10] B. P. Williams, G. Klein, and I. Reid, "Automatic relocation and loop closing for real-time monocular SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1699–1712, 2011. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.41>
- [11] K. Tanaka, "Incremental loop closure verification by guided sampling," *CoRR*, vol. abs/1509.07611, 2015.
- [12] K. Tanaka and E. Kondo, "Incremental RANSAC for online relocation in large dynamic environments," in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, May 15-19, 2006, Orlando, Florida, USA, 2006*, pp. 68–75.
- [13] T. Kanji, "Cross-season place recognition using nbnn scene descriptor," in *Proc. IEEE/RSJ Int. Conf. IROS*. IEEE, 2015.
- [14] A. Masatoshi, C. Yuuto, T. Kanji, and Y. Kentaro, "Leveraging image-based prior in cross-season place recognition," in *ICRA*. IEEE, 2015, pp. 5455–5461.
- [15] T. Kanji, "Self-localization from images with small overlap," in *IEEE/RSJ IROS*, 2016.
- [16] N. Tomomi and T. Kanji, "An incremental scheme for dictionary-based compressive slam," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 872–879.
- [17] M. Brubaker, A. Geiger, R. Urtasun, et al., "Lost! leveraging the crowd for probabilistic visual self-localization," in *Proc. IEEE Conf. CVPR*, 2013, pp. 3057–3064.